Measurement-based Research: Methodology, Experiments, and Tools

Dimitri Papadimitriou
Alcatel-Lucent Bell Labs
Antwerp, Belgium
dimitri.papadimitriou@alcatel-lucent.com

Lluís Fàbrega, Pere Vilà
Universitat de Girona
Girona, Spain
{lluis.fabrega,pere.vila}@udg.edu

Davide Careglio
Universitat Politècnica de Catalunya
Barcelona, Spain
careglio@ac.upc.edu

Piet Demeester
Ghent University
Gent, Belgium
piet.demeester@intec.ugent.be

This article is an editorial note submitted to CCR. It has NOT been peer reviewed. The author takes full responsibility for this article's technical content.

ABSTRACT

In this paper, we report the results of the workshop organized by the FP7 EULER project on measurement-based research and associated methodology, experiments and tools. This workshop aimed at gathering all Future Internet Research and Experimentation (FIRE) experimental research projects under this thematic. Participants were invited to present the usage of measurement techniques in their experiments, their developments on measurement tools, and their foreseeable needs with respect to new domains of research not currently addressed by existing measurement techniques and tools.

Categories and Subject Descriptors

C.2.3 [**Network operations**]: Network monitoring; C.2.5 [**Local and Wide-Area Network**s]: Internet; C.4.2 [**Performance of Systems**]: Measurement techniques|Active

Keywords

Measurement, Methodology, Experimental research, Tools

1. INTRODUCTION

The foundational objectives of Future Internet Research and Experimentation (FIRE) [1] have lead to the inception of experimentally-driven research as a visionary multidisciplinary investigation activity, defining the challenges for and taking advantage of experimental facilities. Such investigation activity would be realized by means of iterative cycles of research, oriented towards the design and large-scale experimentation of new and innovative paradigms for the Internet modeled as a complex distributed system. The refinement of the research directions should be strongly influenced by the data and observations obtained from experiments performed at previous iterations; thus, being "measurement-based", which in turn requires the specification of the relevant criteria and metrics as well as their corresponding measurement tools.

The rationale of FIRE is thus to create a dynamic between elaboration, realization, and validation by means of iterative cycles of experimentation. Its realization was however already less obvious and rapidly confronted to the lack of computer communication/networking experimental model. Moreover, the "validation by experimentation" objective opens a broad spectrum of experimentation methods and tools ranging from simulation to real system/prototype experimentation. The selection of the tool(s) depends on 1) the object of experimentation (referred to as the experimental corpus), 2) the nature and properties of the measurements, and 3) the cost function that itself depends on complexity, experimental and running conditions but also on the level of abstraction (referred to as "realism"). In this context, measurements and tools play a fundamental role in experimental research that aims at the validation by experimental evaluation of project outcomes including protocols, systems and components, etc., by means of reliable and verifiable tools, including on-line analysis of measurement data, mining of data, and diagnostic.

As experimental validation of "elaboration and realization" involves a very broad set of measurement methods and tools, the initial goals of the workshop organized by the FP7 EULER project on May 9, 2012 in Aalborg, Denmark [2] were to collect detailed information on i) the current developments on measurements in experimental research projects within FIRE in well-established research areas, including wireless and sensor networks, routing, etc., and ii) the anticipated needs in new research areas including information-centric networking, programmable components / networks, etc. Based on this input, the objectives of this workshop were the following:

- Identify i) which measurement tools have been developed and applied inside/outside the scope of FIRE experimental research facilities: determine missing elements for large-scale experiments on these facilities, ii) which tools can be combined (with potential extension(s)) to conduct larger experiments, e.g., multiple STREPs joining their efforts, and which tools would be missing to perform larger experiments, and iii) which tools are mature enough to start as basis for re-use by other projects.
- Determine what are the foreseeable needs and their commonality with respect to new domains and areas of research (not currently addressed) by existing measurement techniques and tools.
- Document the lessons learned and best practices in tools development for measurement-based experimental research.
- Identify which initiative(s) can be initiated and realized by means of cooperation between research projects from a directory of tools accessible to the research community at large up to the joint development of measurement tools (under which conditions, etc.)

This paper is organized as follows. Section 2 outlines the measurement-based research methodology including the procedures, objectives, and criteria measurement that tools shall meet to declare scientific validity of the results they produce. The measurement-based experiments conducted in FIRE projects are documented in Section 3. The tools developed for realizing these experiments are detailed in Section 4. Finally, Section 5 summarizes the lessons learned and best practices as recorded during the workshop discussions. This section also provides several key recommendations drawn form this workshop on experimental-based measurement and associated tools.

2. MEASUREMENT-BASED RESEARCH: METHODOLOGY

2.1 Measurement Procedure and Process

Measurement refers to metrology which is defined by the International Bureau of Weights and Measures (BIPM) as "the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology." The term metrology includes all aspects of measurement (theoretical and practical) [3] [4]: starting from the "principles of measurement", which represent the scientific basis for measurement, the "method of measurement" (logical sequence of operations) is instantiated by a measurement (set of operations). The measurement process is instantiated in a measurement procedure having the "measurand" (quantity that is to be measured) as its inputs, the control variables, and the output representing the "measurement results". The measurement process comprises 3 distinct steps: 1) design of a measurement method, 2) application of measurement method rules, and 3) analysis of the measurement results. To carry out a measurement task, an experimenter should design and execute a measurement procedure (corresponding to the measurement function μ) which consists of a set of operations, specifically described, for the performance of a particular measurement according to a given measurement method. Note that the results of the measurement can be influenced by external quantity during the measurement process.

2.2 Measurements Objectives

As documented in [5], measurements aim at determining not only i) the value of a quantity but also at determining, ii) the distributions of quantities in time, in space, and in time and space, iii) the mathematical representations of quantities or their distributions, iv) the relations between quantities, their distributions or representations, and v) the parameters of such relations. The results of measurements of types (i) and (v) are expressed in terms of numbers. The results of measurements of types (ii), (iii) and (iv) may have the form of numbers, series of numbers, functions or series of functions -given in tables-, or analytically. When measurements of type (v) are considered, then the parameters of relations between quantities are often treated as new quantities (e.g., resistance, inductance, capacitance), but the diversity of the investigated relations (e.g., non-linearity, dependence on frequency) breaks the quantitative concept of measurement.

From this perspective, measurement-based experimental research aims at complementing the rigorous performance analysis and simulation-based evaluation. The results are more realistic and can contribute to validate and to fine tune the execution of algorithms. A large variety of realistic topologies, mobility profiles, and traffic patterns is required. Novel network parameters as well as performance monitoring measures (and their trade-offs) arise. Ad hoc approaches are useful but there is a need to converge to widely accepted, common integrated measurement methods, systems and tools.

2.3 Measurements Properties and Criteria

On the other hand, measurement results obtained by means of experiments have to verify certain properties and criteria in order to accept experimental research as a viable mean to declare scientific validity of the results these experiments produce. These properties and criteria mainly include reliability, repeatability, reproducibility, and verifiability. In turn, they constraint the experimental corpus, methodology, and determine the properties and criteria that shall be met by measurement tools.

2.3.1 Reliability

Reliability is defined as the probability that the measurement function μ performs its intended measures (output) during a specified period of time under stated conditions. More formally, referring to Fig.1, where the experimental corpus is modeled by a function F with input vector x and output y = F(x), reliability is verified when $\exists\, [t_1,t_n]$ and $\varepsilon << 0$ such that $\forall\, k \in N,\ 1 \leq k \leq n$, $\mu(y(t_k)) = \mu(F(x(t_k))) \wedge y(t_k) \in [\mu(y(t_k))-\varepsilon, \mu(y(t_k))+\varepsilon]$, where $y(t_k) = F(x(t_k))$. Reliability implies as a minimum requirement that the components of the experimental corpus remain operational (i.e., do not fail or halt) during this time period. Furthermore, measurement results are reliable if they remain consistent (within a certain well-defined range) during that period.

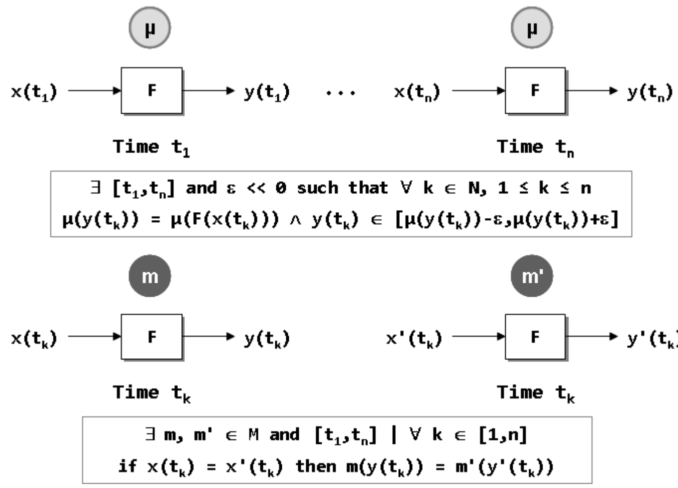

Fig.1: Reliability

In order to assert the reliability of a given measurement tool implementation m ($\in$ M $\equiv$ measurement program set) of a given measurement function μ, it is common to compare the results of measurements produced by m with those obtained for the same time period by means of another implementation m' ($\in$ M) of the same function μ. Referring to Fig.1, $\exists$ m, m' $\in$ M and $[t_1,t_n]$ such that $\forall\, k \in [1,n]$, if $x(t_k) = x'(t_k)$ then $m(y(t_k)) = m'(y'(t_k))$.

2.3.2 Repeatability

Repeatability is a temporal criterion associated to measurement results. This term is used when multiple execution of a given experiment (repetition) using the same configuration, running conditions, and input yields the same output. Correct experimental method and usage of models, execution of algorithms, and output processing is required in order to guarantee the repeatability of measurement results. More formally, referring to Fig.2, repeatability is verified when the following condition is met $\forall\, k \in N,\ k \geq 1$, if $x(t_k) = x(t_{k-1})$ then $\mu(y(t_k)) = \mu(y(t_{k-1}))$, where $y(t_k) = F(x(t_k))$ and $y(t_{k-1}) = F(x(t_{k-1}))$.

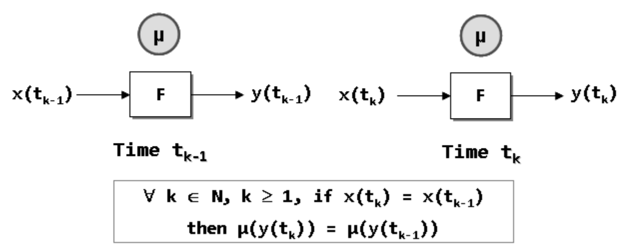

Fig.2: Repeatability

2.3.3 Reproducibility

Reproducibility is a spatial criterion associated to measurement results that can be obtained when a given experiment performed on a given experimental system u ($\in S \equiv$ experimental system set) is replicated over a similar but different experimental system v ($\in$ S). This can mean different experimental platform/facility, operating system, etc. Typically, reproducibility comes into play when a third party performs the same experiment to determine the scientific validity of the output of an experiment. More formally, referring to Fig.3, reproducibility is achieved when $\exists$ u, v $\in$ S such that if the input vector $x_u = x_v$ then $\mu(y_u) = \mu(y_v)$, where $y_u = F(x_u)$ and $y_v = F'(x_v)$.

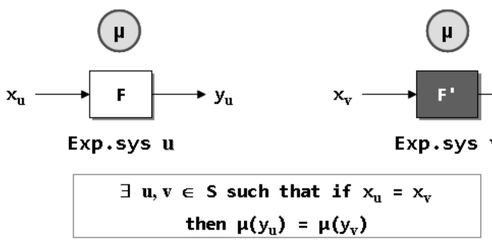

Fig.3: Reproducibility

2.3.4 Verifiability

The results of experimental measurements are verifiable if the output of the experimental corpus modeled by the function F: $x(t) \rightarrow y(t) = F(x(t))$ can be confirmed against a formal model H: $x(t) \rightarrow H(x(t))$; implying that the measurement results shall comply with the output of the model H (output described as a function of the input vector and the experimental parameters). Referring to Fig.4, measurement results are verifiable if there exists a formal model H: $\mathfrak{R}^n \rightarrow \mathfrak{R}$: $x(t) \rightarrow H(x(t))$ and $\varepsilon << 0$ such that at time t_k, $H(x(t_k)) \in [\mu(y(t_k))-\varepsilon;\mu(y(t_k))+\varepsilon]$, where $\mu(y(t_k)) = \mu(F(x(t_k)))$. One often considers that verifiability is achieved by comparing the results of an experimental measurement against a reference system RS (assumed as representative of the real system): $\exists$ u $\in$ S and $\varepsilon << 0$ such that if $x_u = x_{rs}$ then $y_{rs} \in [\mu(y_u)-\varepsilon;\mu(y_u)+\varepsilon]$.

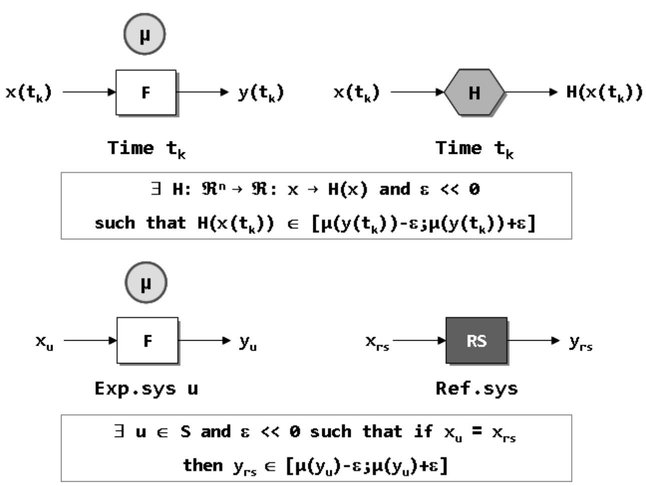

Fig.4: Verifiability

Achieving verifiability for a representative sample (to avoid sampling bias) of unbiased measurement results whose size is determined so as to reduce the sampling error (and satisfy a given confidence interval and level given the finite but often large number of available results) enables in turn to generalize the conclusion(s) that can be drawn from experimental measurements.

3. MEASUREMENT-BASED EXPERIMENTS IN FIRE

This section describes the experiments (per research area) as performed within FIRE that are relying on experimental measurements.

3.1 Large-Scale Experiments

The deployment of experimental facilities is a crucial requirement for validating Internet research activities, and many efforts are being done to achieve it. The federation of testbeds, each of them addressing different applications or technologies, offers a richer and more powerful experimental facility to enable heterogeneous and large-scale Internet-oriented research. However most deployed testbeds have been built with a certain application scope in mind and usually uses its own control framework.

The OpenLab project [6] brings together a number of different and diverse testbeds, such as wireline (PlanetLab Europe), wireless (NITOS, w-iLab.t), multimedia (WIT IMS), high precision measurement (ETOMIC), or emulation testbeds (HEN), which use different control framework and tools (e.g., MySlice, Federation Computing Interface and Federation Scenario Description Language, cOntrol and Management Framework and OMF Experiment Controller, or Network Experimentation Programming Interface). The main goal of the OpenLab project is to enable transparent access to combinations of resources from different testbeds, addressing the interoperability challenges at several levels by i) using tools tailored for a given testbed in other testbeds, ii) migrating experiments performed in old testbeds to new ones, iii) reproducing experiments in similar, yet different testbeds, and iv) extending experiments to enlarged scale or enhancing experiment to a broader scope. Examples of experiments being conducted in OpenLab include: mechanism to find the location of the content servers in Content Delivery Networks (CDNs), replica placement and virtual resource mapping framework for realizing a large-scale scenario of wireless CDN (i.e., CDNs where some clients use wireless access links), and an architecture based on locator-identifier separation for supporting mobility and multi-homing in the Internet.

The NOVI project [7] aims at providing control, management and monitoring planes to allow the federation of various virtualized testbeds, consisting of heterogeneous resources, enabling experimenters to request, reserve, use and update a great deal of virtualized resources in the federation, tailored to their needs. The necessary abstraction of the managed entities is provided by information models, which should support virtualization concepts, vendor independence (of the physical resources), monitoring and measurement concepts and management policies. The information model facilitates the control and management of the individual platforms, and the communication between them.

3.2 Wireless and Sensor Networks

In general, the experiments conducted in this research area share the following characteristics:

- It is difficult to reproduce and validate the experiments due to the impossibility to isolate them from radio interferences coming from other sources. Also the energy conditions are never exactly the same because of the conditions and life of batteries.
- Experiments make use of highly heterogeneous equipment including different types of sensors (temperature, humidity,

cameras, etc), actuators and different radio technologies (Wi-Fi/802.11, WiMax/802.16, 3G, 4G, ZigBee, Bluetooth, etc.).

- Dynamicity and mobility have to be included, e.g., wireless devices can be randomly distributed over an area or placed regularly, can also be added or removed randomly while the experiment takes place, and can also be mobile following specific patterns or moving randomly.

For spectrum sensing through heterogeneous devices, the CREW project [8] proposes a benchmarking framework that comprises the following steps:

- Pre-calibration of the heterogeneous hardware, using a common metrics. The main metric is the Power Spectrum Density (PSD), with specific values of the span, resolution bandwidth, sweep time, time resolution, and energy detection threshold. Device heterogeneity induces different signal strength attenuation in the receiver chain, called power offsets, which have to be measured for later processing of the measured data (in the post processing stage).

- Set up the experiment, using metadata to unambiguously describe the experiment. Metadata could include the following: transmission signal pattern, transmission power level, device name, device location, device power offset, start time, frequency bins, resolution bandwidth, etc.

- Measure, using a clear methodology. The target would be the characterization and comparison of heterogeneous spectrum sensing devices. The approach would be to select a frequency band, to configure the transmission signal and to measure the PSD.

- Comparison, using a common data format, post processing and scores. Measurements would be in the form of a matrix containing PSD and relative time stamps. The post-processing would compensate for hardware heterogeneity, through pre-calibrated power offsets, and for software heterogeneity, through averaging and re-sampling the PSD matrix so that all devices have a common reporting rate in time domain. Then, through a given energy detection threshold, signal detection is determined. Finally, a “device performance score”, which indicates its performance, is derived.

Note that a "metric" here means a quantitative measure of a certain quality of the System Under Test (SUT), while "score" means an abstraction of a set of metrics (e.g., device performance, reliability of the experiment, cost, repeatability). Score hides performance evaluation details and is useful when comparing a large number of experiments, to evaluate a solution by non-experts, and to automate the performance evaluation.

The wireless NITOS testbed developed by the CONECT project [9] allows experimental work both at the packet level and at the MAC layer so that researchers can analyze and implement new cooperative schemes. The testbed includes a large number of heterogeneous wireless devices including Wi-Fi nodes, USRP boards, sensors, WiMax nodes, 4G and 3G femtocell components, which allows for a wide range of experiments involving different wireless technologies. Access to resources takes place through the slice abstraction, isolated resource containers accessible by one or multiple users. Apart from the wireless network itself, the testbed comprises three discrete wired local networks: the control network to log into the nodes via a server; the chassis network to power nodes on or off; and an OpenFlow-capable experimental network. The testbed also provides tools to assist experimenters in assessing the testbed’s wireless environment properties and selecting an appropriate topology for their experiment. Examples of experiments conducted in the NITOS testbed include:

- Design and evaluation of cooperative networks, which exploit different paths through the aid of possible relays that carry out the traffic. The goal is to increase throughput and minimize power consumption.

- Demonstration of a scenario where a vehicle equipped with sensors gathers measurements from its environment and communicates opportunistically with road-side units to forward the measurements to a centralized framework for their storage and analysis.

The main objective of the HOBNET project [10] is to ease the development of applications for automation and energy-efficiency of smart/green buildings through sensor networks. For this purpose, several testbeds are available, with high diversity of devices (sensors actuators, etc.) and different thematic emphasis (energy, tracking, visualization, etc.). Many different applications have been evaluated on these testbeds such local adaptation to presence, CO_2 monitoring, garden watering, maintenance control, or electric device monitoring. Several critical options for the experimental scenario have been detected: structured topologies versus randomized deployments; homogeneous sensor deployments versus heterogeneous deployments; all sensors running at the start of the experiment versus sensors added during network evolution; uniform node density versus high density diversity; static deployments versus mobile deployments (and hybrid combinations). Performance evaluation experiments aim at determining the scalability with respect to the network size, the fault-tolerance properties (this implies the need for diverse fault models), and inherent trade-offs such as energy versus time.

3.3 Routing

In the context of FIRE, Information-Centric Networking (ICN) and inter-domain Internet routing at large-scale are two research areas where routing is being experimentally investigated.

ICN is a novel paradigm where the network layer provides content to applications, instead of providing connectivity between hosts/terminals. The base functionality of ICN relies on i) content addressing through a scheme based on names or identifiers (that do not include references to their location); and ii) routing content requests toward the closest copy of the content with such a name (name-based anycast routing). The CONVERGENCE project [11] develops the ICN functionality as part of the network layer of its architecture, which aims at enhancing the Internet with a content-centric, publish-subscribe service model. The experimental network protocols to be implemented would run on some dedicated routers. Once an end node sends a contents request, border routers use a routing table based on content names to forward the request to the next node, till the serving node, which then sends the content back. Besides routing by name, border routers also perform caching, and in the name-based routing table also store entries related to the cached content. Therefore, routing tables in nodes include IP routes, name-based routes and cached content index. Routing tables are not complete, and in case of a missing routing entry, a centralized routing engine called Name Routing System (NRS) provides the entry, which is temporary stored in the table. Experiments in ICN have been conducted to verify that current technology scales in terms of memory size of the local routing tables and supports the route lookup rate required at the NRS node. Other experiments have been performed on the

PlanetLab Europe network testbed [6] with a scenario of 20 nodes (one acting as NRS node), to evaluate the performance of the routing-by-name functionality. Download time, routing table size, number of route lookups at the NRS node, cache size and the amount of protocol messages exchanged by nodes were some of the measured quantities. In all these experiments, the ICN is used to distribute current web contents.

The Internet at the inter-domain scale is another area of research on routing in FIRE projects. It has been long recognized by the scientific community that the most fundamental problems faced by the Internet architecture are the scalability, convergence, and stability properties of its inter-domain routing system based on the Border Gateway Protocol (BGP). Solving these problems requires addressing multiple dimensions altogether: i) the routing table size growth resulting from an increasing number of routing entries, and ii) the routing system dynamics characterized by the routing information exchanges produced by topological or policy changes. Research on new paradigms for distributed and dynamic routing schemes suitable for the Internet and its evolution is the main goal of the EULER project [12]. Investigation is going in two directions: compact routing schemes, which omit some topology details to achieve a good tradeoff between reducing the routing table size and the increase of the routing path length; and geometric routing schemes, where nodes are assigned virtual coordinates in a metric space and the next node along the routing path is the neighbor that is nearest to the destination in tat space. However, experimentation of new routing schemes for the Internet is confronted to the situation where none of the existing facilities recreates the actual running conditions at the scale of the Internet, where these routing schemes can be evaluated before being deployed. As the Internet scale is difficult to reproduce and the routing states and dynamics difficult to model, a new experimental approach is considered: simplifying the actual experimental corpus through functional abstraction, and reproducing significant phenomena through patterns derived from measurements of the actual environment. Therefore the experimental approach followed by the EULER project is the following: develop measurement tools in order to derive representative running conditions of the actual environment, i.e., the Internet topology and its dynamics, to derive patterns for modeling the topology dynamics and then use these (dynamic) models for generating experimental scenarios to be executed in large scale routing scheme simulation and emulation. Examples include here the modeling of the BGP routing paths stability and its evolution over time, and the evolution of the nodal degree.

3.4 Software-Defined Networks

The pan-European research testbed OFELIA [13] makes use of OpenFlow. This technology enables the separation of control, forwarding and processing of data by defining the interactions and the operations performed from a (non-)co-located control element (the controller) to a data plane element (switch, forwarder). The header of an incoming packet is compared to a set of defined "matches", i.e., patterns containing specific values of packet fields (Ethernet, IP, transport), and then some "actions" over the packet are performed (forward, rewrite fields, etc.). The set of packets seen by the switch since the rule was installed defines a flow.

OFELIA project targets to build (and in the second phase interconnect) a set of campus installations (islands) that consist of GNU/Linux-based virtual machines (VMs) interconnected by OpenFlow switches, and to make these facilities available to all researchers. Internally, the physical network consists of two networks, the control network, which provides access to the control interfaces of VMs and to the switch controllers, and the experimental network, which connects data interfaces of VMs and switch ports. Experiments run in "network slices", i.e., virtual networks that share the same physical network, built using FlowVisor, a network virtualization layer. FlowVisor, a special purpose OpenFlow controller, acts as a transparent proxy between OpenFlow switches and multiple OpenFlow controllers. It creates rich slices of network resources, delegates control of each slice to a different controller and enforces isolation between each slice, i.e., one slice cannot affect another's traffic. In OFELIA the experimenter can get access to a testbed island through an OpenVPN tunnel; then he can generate on demand the entire network slice (using Expedient, a special web-based resource allocation tool), access to the VMs through Secure Shell (SSH), and setup the experiment. OFELIA provides also a set of tools to test various aspects of OpenFlow switches and controllers, such as OFLOPS, cbench and OFTest. It also provides a set of traffic generation and measurement tools to perform the experiments.

4. MEASUREMENT TOOLS DEVELOPED IN FIRE

This section describes the different tools developed by the FIRE projects to perform the measurements required by the various experiments documented in Section 3.

4.1 Large-Scale Experiments

Experiments require dedicated tools to generate traffic load (sources, sinks, well-defined flows), validate functionality (protocol message format and sequencing) and evaluate performance (packet delay, jitter, packet loss, link usage, throughput, etc.).

In OFELIA testbeds [13], the experimenter is provided with a built-in set of tools as well as the capacity to install its own external tools. Provided tools are open source (e.g., Wireshark, iperf, etc.) and also high performance test systems (e.g., IXIA T1600 and associated software), which allow to perform customized and predefined tests but also inject and measure traffic at any point of the network.

In the federation of multiple testbeds, the presence of different and diverse measurement tools is a problem. Differences may appear in naming, data representations, units, metadata and data merge, and therefore their integration becomes necessary. The OpenLab [6] solution to this problem deals with the semantics of the information, unambiguously specifying the set of concepts that compose a measurement. The solution comprises three steps: the agreement on a common ontology for network measurements, the definition of mappings between each particular scheme and the common ontology, and the definition of a semantic interface (based on the ontology) able to receive a query from a user and distribute it among all particular measurement repositories. Such ontology is currently being standardized at the European Telecommunications Standards Institute (ETSI) and several mappings are being defined.

A monitoring and measurement framework to allow the federation of virtualized testbeds is proposed by the NOVI project [7]. With the use of a specific monitoring ontology, a wide range of monitoring tools, metrics and databases can easily be integrated in this framework. An experimenter only needs to know the metrics to be measured (e.g., throughput, one-way/two-way delay) independently of the tools installed in the different testbeds and the monitoring service will ensure the proper mapping between the requested metrics and the available tools automatically.

4.2 Wireless and Sensor Networks

The HOBNET project [10] has developed the REST architecture for automation and energy-efficiency of smart/green buildings through sensor networks. In this architecture, every distinguishable and addressable entity is defined a resource (anything from a physical device, like a sensor/actuator, to a web site, XML file, etc.) uniquely identified by an URI. The underlying IPv6/6LoWPAN infrastructure level makes use of IPv6 to integrate heterogeneous technology (sensors, actuators, mobile devices, etc). At a second level of the architecture, several algorithmic models and solutions for smart buildings has been proposed with special care for scalability. On top of this structure, an interface is available for rapid development of building management applications. Then, the proposed experiments can be organically evaluated in the context of the platform integration. One of the modules included in this architecture is the MeasurementsLogger component, whose role is to setup the parameters of the experiments (event generation rate, energy sampling rate, duration, etc) and to monitor the evolution of the evaluated scheme by enabling the logging of the performance measurements (delivery latency, the average energy consumption, and the success ratio).

The wireless NITOS testbed [9] handles measurements using the cOntrol and Management Framework (OMF) adopted from the OpenLab project [6]. The OMF framework enables an efficient management of the heterogeneous resources of this testbed, providing a clear and easy way to define experiments, execute them and collect the results. The OMF Measurement Library (OML) is based on customizable measurement points inside applications running on the resources and provides a well-structured solution for capturing, processing, filtering, storing and visualizing measurements. An OML server is responsible of gathering the measurements and storing them in a database while OML clients are capable of injecting measurements generated at measurement points into streams towards the OML server. Extra features include the OMF graph generator (displaying results) and the proxy OML servers to cover disconnected parts of the same experiment thus enabling mobility support.

4.3 Routing

In experimental research on routing, measurements are performed in the current Internet, and results processed so as to model some aspects useful for the design and the subsequent evaluation of new routing schemes, and in the experimental testbeds where the new algorithms and protocols are implemented in order to evaluate their performance.

In particular, some of the experiments in ICN [11] consider the scenario where ICN is used to distribute current web contents. Therefore deriving ICN traffic traces between web clients and servers from Internet measurements is necessary. A tool has been developed to convert Internet web traffic traces into ICN traces: TCP messages are mapped to ICN messages and, since web server's IP addresses in Internet traces are usually anonymous, they are randomly assigned to a set of public IP addresses of the most used domain-names.

Experimental scenarios for Internet routing schemes require models of the Internet topology and its dynamics, derived from measurements. The EULER project [12] has developed tools to measure the node degree distribution, the dynamics of the network topology and events, and the stability of BGP routing paths:

- Distributed UDP Ping, a tool for measuring the node degree. UDP Ping works as follows: the application sends a UDP packet towards an unallocated UDP port of a target router; the router sends back an ICMP (destination unreachable) packet through one of its interfaces; from the received packet the application finds out the IP address of the target router. Distributed UDP ping comprises a large set of monitors located in different places that together hopefully obtain the whole set of IP addresses of the target router, and then estimates its number of interfaces. From the data obtained the degree distribution can be estimated. This tool has been deployed and experimented on PlanetLab [6].
- Tracetree - radar, a tool for measuring topology changes, routing path dynamics, and load balancing. Tracetree is similar to traceroute but it obtains a routing tree of IP paths from 1 source to many destinations. Radar performs periodic measures with Tracetree to observe the dynamics. Dynamics in routing paths come from physical topology changes, routing path dynamics, and load balancing. From the data obtained a dynamics topology model that fits the observations can be derived. This tool has been also deployed and experimented on PlanetLab [6].
- A tool to measure the stability of BGP routing paths. BGP routes are affected by two types of instability, policy-induced (conflicting policy interactions) and protocol-induced (path exploration), which may lead to non-deterministic unstable states and delayed BGP convergence. A set of stability metrics that measure the variation of the BGP routing table entries at periodic time intervals enables to determine the stability of its routing paths. For this purpose, a dedicated tool has been developed that parses significant volumes of real BGP datasets obtained from collectors under the control of RouteViews [13] and computes the stability metrics.

Concerning the performance evaluation of new routing paradigms in experimental testbeds, common tools can be used to measure the interesting quantities: web download time, routing table size, cache size, number of route lookups, or number of protocol messages in [11]; routing table size, number of routing messages, convergence time or routing path stretch in [12].

5. BRIDGING THE GAP

This (concluding) section aims at documenting the lessons learned and best practices as recorded during the workshop discussions and the panel session conducted at the end of the workshop. It also provides several key recommendations drawn from the outcomes of the workshop on experimental-based measurement and associated tools.

5.1 Lessons Learned and Best Practices

The following observations can be drawn from the discussions held during the workshop:

- The development of dedicated measurement tools is time-consuming (to comply with the verification, reliability properties as documented in Section 2) and existing tools when re-used (and thus re-usable) provide relatively limited extensibility potential.
- Testbeds are of different nature (wired, wireless, different hardware, different set-ups, etc.) and experiments themselves are conducted for different measurement purposes (even when performed on the same testbed); henceforth, measurement tools designed for experiments to be realized on a given testbed are likely to be incompatible with each other; thus measurements are often not reproducible.

Moreover, the reliability of measurements is more challenging to achieve in open testbeds where all running conditions are not under the control of the experimenter.

- Experiments conducted in wireless and sensor environments are repeatable but only up to a certain probability that the measurement tool performs its intended measure (output) during a specified period of time under stated conditions. Ensuring these conditions are verified is challenging in open testbeds/experimental environments.
- Few (if not none) projects dedicate effort to verify their measurement results. The reasons are multifold: complexity of the experiments and variety of components they involve (thus their modeling), unavailability of comparable real system (since the experimented corpus is by nature unavailable), time required to perform systematic verification, etc. The measurement verification phase is often limited to ad-hoc (or eventually more systematic) comparison with similar experiments acting either as experimental model or reference system. For this purpose, measurement tools should also work / be adapted (from the testbed/prototype experiment) to also run in the context of emulation and simulation experiments. Such practice would facilitate the comparison of results.
- There is an increasing need for specifying well-defined interfaces between different measurement tools with implementations adapted to different equipment, for standardizing the formats of the collected data and, if possible, also the control of the experiment. Moreover, the factorization of the code by means of well defined and well documented software modules shall be put into practice so that measurement tools developers could use these software modules independently and in turn improve their reusability.
- It is also recognized that the advent of new (or rejuvenated) technology research areas including programmable/software-defined networks induce specific needs such as on-line code verification and software robustness testing.

5.2 Recommendations

All workshop participants agreed that initiating right away a dedicated initiative would be advisable on measurement tools, testbeds and experimental measurement-based research in networking/communication technologies in order to communalize and to factorize the development of measurement tools. Moreover, this initiative should remain in charge of the research community to avoid introduction of proprietary measurement software. Here below, we summarize the recommendations that can be derived from the workshop:

- Concerning the question "How can the huge number of tools that already exist be shared ?" the research community should create joint working groups involving different research projects and identify common developments that can be shared with other projects. At the end of the development and validation phase, tools should be made available as open-source code to the research community at large (by means of openly and easily accessible repository).
- On the other hand, bring measurement tools developers from different research projects is recommended to share their experience and work together in order to progressively specify a common modular baseline when developing measurement tools together with common data formats and generic interfaces (including the control interfaces). Moreover, all measurement tools and associated standards should be open-source.
- The implementation of a measurement tool repository should start with a simple and easily accessible repository with progressive addition of improvements (bottom-up) rather than designing a powerful but empty system (top-down). Even further, building such repository could start by just providing data from experiments (data sets) so that other researchers could re-use these data to further analyze them and obtain additional results. It is also important that all these tools could also work / be adapted from the testbed scenario to the emulation scenario and to the simulation scenario. This will facilitate the comparison of results.
- Finally, participants also recommended, as the amount of collected data increases (due to the scale of experiments), that the development of measurement tools shall not be limited to actual measures and their collection but also to data analytics and related tools.

ACKNOWLEDGMENTS

The "Measurement-based research: methodology, experiments and tools" workshop has been organized by the FP7 EULER project (Grant no.258307) funded by the European Commission.

6. REFERENCES

[1] FIRE EC Website: *http://cordis.europa.eu/fp7/ict/fire/*

[2] Workshop on "Measurement-based research: methodology, experiments and tools", Aalborg, Denmark, May 9, 2012. Abstracts and presentations available at *http://www.euler-fire-project.eu/*

[3] International Organization for Standardization (ISO), International vocabulary of metrology - Basic and general concepts and associated terms (VIM), ISO/IEC Guide 99:2007.

[4] A.Abran and A.Sellami, Initial Modeling of the Measurement Concepts in the ISO Vocabulary of Terms in Metrology, Software Measurement and Estimation - Proceedings of 12th International Workshop on Software Measurement (IWSM), Magdeburg, Germany, Oct. 7-9 2002, Shaker-Verlag, Aachen, 2002.

[5] A.J.Fiok, J.M.Jaworski, R.Z.Morawski, J.S.Oledzki and A.C.Urban, Theory of measurement in teaching metrology in engineering faculties, Warsaw University of Technology, PL-00-661 Warsaw, Poland.

[6] OpenLab FP7 project website: *http://www.ict-openlab.eu/*

[7] NOVI FP7 project website: *http://www.fp7-novi.eu/*

[8] CREW FP7 project website: *http://www.crew-project.eu/*

[9] CONECT FP7 project website: *http://www.conect-ict.eu/*

[10] HOBNET FP7 project website: *http://www.hobnet-project.eu/*

[11] CONVERGENCE FP7 project website: *http://www.ict-convergence.eu/*

[12] EULER FP7 project website: *http://www.euler-fire-project.eu/*

[13] OFELIA FP7 project website: *http://www.fp7-ofelia.eu/*

[14] University of Oregon Route Views project website: *http://www.routeviews.org*